

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: May 29, 2002, 11:38:48 ; Search time 34.14 Seconds  
(without alignments)  
3278.492 Million cell updates/sec

Title: US-08-485-355b-50  
Perfect score: 3374  
Sequence: 1 MGDAGVASORPHNRGTRNV.....GKIAHVARARRARRAARAN 647

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_virus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3374	100.0	647	12	082462 helicoverpa
2	2205	65.4	644	12	090063 nudaurelia
3	435.5	12.9	757	12	09YK87 thossea asig
4	386.5	11.5	612	12	09YRB2 nudaurelia
5	134	4.0	2147	2	09L950 pseudomonas
6	132	3.9	886	12	09QPB7 human herpe
7	131.5	3.9	856	12	056854 human herpe
8	130	3.9	597	16	09RY11 deinococcus
9	127	3.8	466	5	095S10 drosophila
10	127	3.8	1212	5	095T50 drosophila
11	127	3.8	2016	5	09V439 drosophila
12	127	3.8	2016	4	09NBA1 drosophila
13	126	3.7	3570	4	099552 homo sapien
14	125.5	3.7	717	2	09KHB3 bacillus ci
15	125.5	3.7	1269	17	097Z06 sulfolobus
16	125	3.7	2314	2	069822 streptomyces

17	123.5	3.7	1204	17	09UY51 pyrococcus
18	122	3.6	1314	4	09UMZ2 homo sapien
19	121.5	3.6	699	2	048494 kurtzia zop
20	120.5	3.6	769	12	09PZT4 human parvo
21	120.5	3.6	781	12	P83318 human parvo
22	120.5	3.6	979	2	09XAS7 streptococ
23	120	3.6	873	16	09RX36 delnoco
24	119.5	3.5	851	12	09OP86 human herpe
25	119.5	3.5	889	4	014917 homo sapien
26	119.5	3.5	907	12	066537 human herpe
27	119.5	3.5	1425	5	09VIT9 drosophila
28	119.5	3.5	1293	16	099QY4 staphylococ
29	119	3.5	2271	16	092PFG5 listeria in
30	119	3.5	2809	5	061230 lytechinus
31	118.5	3.5	3029	16	055582 syncocyst
32	117.5	3.5	781	12	09UGP8 human parvo
33	117.5	3.5	884	16	09RMT0 delnoco
34	117.5	3.5	26926	4	010466 homo sapien
35	116	3.4	4578	16	050379 mycobacteri
36	116	3.4	1496	5	042181 fuqu rubrip
37	115.5	3.4	1496	5	09NFV5 drosophila
38	114.5	3.4	781	12	065789 human parvo
39	114.5	3.4	828	10	P89316 human parvo
40	114	3.4	658	10	004238 vicia faba
41	113.5	3.4	658	12	066536 human herpe
42	113.5	3.4	781	12	085117 human parvo
43	113.5	3.4	781	12	090200 human parvo
44	113.5	3.4	781	12	P90221 human parvo
45	113.5	3.4	781	12	P90222 human parvo

## ALIGNMENTS

RESULT 1  
ID ID 082462 PRELIMINARY: PRT: 647 AA.  
AC 082462:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE COAT PROTEIN.  
GN P71.  
OS Helicoverpa armigera stunt virus.  
OC Viruses: ssRNA positive-strand viruses, no DNA stage: Tetraviridae;  
OC unclassified Tetraviridae.  
OX NCBI\_TaxID=37206;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BLACK MOUNTAIN;  
RX MEDLINE=97201539; PubMed=9049325;  
RA Hanzlik T.N., Dorian S.J., Johnson K.N., Brooks E.M., Gordon K.H.;  
RT "Sequence of RNA2 of the Helicoverpa armigera stunt virus  
(Tetraviridae) and bacterial expression of its genes.";  
RL J. Gen. Virol. 76:799-811(1995).  
DR EMBL: L37299; AAC37885.1; -;  
DR MEROPS: A21.001; -;  
SQ SEQUENCE 647 AA: 70670 MW: 08205B8CB53CFBD2 CRC64;

Query Match 100.0%; Score 3374; DB 12; Length 647;  
Best Local Similarity 100.0%; Pred. No. 1e-230;  
Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGDAGVASORPHNRGTRNVASNTVYVNGRRNORRTGROVSPDNFTAAADLAOST 60  
DB 1 MGDAGVASORPHNRGTRNVASNTVYVNGRRNORRTGROVSPDNFTAAADLAOST 60  
QY 61 DANTYTPANTISSMEPEFNMAKGRIDLDSDSIGWFFKYLDPAGATESARAAGEVSKIIPDG 120  
DB 61 DANTYTPANTISSMEPEFNMAKGRIDLDSDSIGWFFKYLDPAGATESARAAGEVSKIIPDG 120  
QY 121 LVKESVDAEIREIYNECPVTVDSVPLDGRWLSLIFSFPNFTAVVAANVENKMSL 180

```
Db 121 LVFSDAEIREIYNECPVYTDVSVLDCRQMSLSIFSPMFRTAYVAANENKMSL 180
Qy 181 DVNDLLEIMNLADMRVYVDSQWIMFTNDTTYVYRIRLRLRPYDVPDEGLVPTVSD 240
Db 181 DVNDLLEIMNLADMRVYVDSQWIMFTNDTTYVYRIRLRLRPYDVPDEGLVPTVSD 240
Qy 241 YRLTYKATCEANMPTLVDOGFWMIGGOYALPTSLPODYVSEAYALHTLFARPPSSAAL 300
Db 241 YRLTYKATCEANMPTLVDOGFWMIGGOYALPTSLPODYVSEAYALHTLFARPPSSAAL 300
Qy 301 AFWAGIAPGCGTAPAGTAPAEQASGGYLTWRHNGTTFPAGSVSYLPEGFALERYDND 360
Db 301 AFWAGIAPGCGTAPAGTAPAEQASGGYLTWRHNGTTFPAGSVSYLPEGFALERYDND 360
Qy 361 GSWTDFASAGDTVFRQVAADVEVYVNNPAGGSAPEFTVRPSPNATYTVRNTLLET 420
Db 361 GSWTDFASAGDTVFRQVAADVEVYVNNPAGGSAPEFTVRPSPNATYTVRNTLLET 420
Qy 421 RPSRRLELPMPPADFGQTVANNPKIEOSLKEITLGCYLVHSMKRNPFOLTPASSFGAV 480
Db 421 RPSRRLELPMPPADFGQTVANNPKIEOSLKEITLGCYLVHSMKRNPFOLTPASSFGAV 480
Qy 481 SFNNPGYERTRLPDYTGIRDSFDQNMSTAVAHFRSLSHSCSIYTKYQGWEGVTNVP 540
Db 481 SFNNPGYERTRLPDYTGIRDSFDQNMSTAVAHFRSLSHSCSIYTKYQGWEGVTNVP 540
Qy 541 FQGFAGHGLKNEIELCLADDLATRLTGYVPATDNFAAASAFANMLSSVLKSEATSI 600
Db 541 FQGFAGHGLKNEIELCLADDLATRLTGYVPATDNFAAASAFANMLSSVLKSEATSI 600
Qy 601 IKSVGETAVGAOSGLAKLPGLIMSVPGKTAARVARRARRARRARAN 647
Db 601 IKSVGETAVGAOSGLAKLPGLIMSVPGKTAARVARRARRARRARAN 647
```

```
RESULT 2
Q90063 PRELIMINARY: PRT: 644 AA.
ID 090063
AC 090063;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CAPSID PROTEIN.
GN CAPSID PROTEIN, CP.
OS Nudaurelia capensis omega virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
OC Omegatetravirus.
OX NCB1_Taxid=12541;
RN RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92391097; PubMed=1519360;
RA Agrawal D.K., Johnson J.E.;
RT "Sequence and analysis of the capsid protein of Nudaurelia capensis
omega virus: an insect virus with T = 4 icosahedral symmetry.";
RL Virology 150:806-814(1992).
DR EMBL; S43937; AAB23198.1; -.
DR MEROPS: A21.001; -.
SQ SEQUENCE 644 AA: 69863 MW: EAA1C9AB50F24828 CRC64;
```

Query Match 65.4%; Score 2205; DB 12; Length 644;  
Best Local Similarity 65.8%; Pred. No. 7.1e-148;  
Matches 427; Conservative 69; Mismatches 135; Indels 18; Gaps 6;

```
Qy 8 SQPHRRGRNRYVSNANTYVNGRRNOR--RRTGROVSPDNFTAAADLQSLDANTY 65
Db 3 SNSASGRSRNRVRIANTYVNAVKORQARRRARRANNIDVNTAAAGLQSLDANTY 62
Qy 66 TEPANISMPDEFNRMAKGIKIDLSDSIGWYFKYLDPAGETESARAAGEYSKIPDGLVKS 125
Db 63 TTFPTNATMPDEFNRMAKGIKIDLSDSIGWYFKYLDPAGETESARAAGEYSKIPDGLVKS 122
```

```
Qy 126 VDAEIREIYNECPVYTDVSVPLDGRQMSLSIFSPMFRTAYVAANENKMSL 185
Db 123 VDAEIREIYNECPVYTDVSVPLDGRQMSLSIFSPMFRTAYVAANENKMSL 185
Qy 186 LIEMNLADMRVYVDSQWIMFTNDTTYVYRIRLRLRPYDVPDEGLVPTVSD 245
Db 183 LIEMNLADMRVYVDSQWIMFTNDTTYVYRIRLRLRPYDVPDEGLVPTVSD 242
Qy 246 KATICEANMPTLVDOGFWMIGGOYALPTSLPODYVSEAYALHTLFARPPSSAALAFVNA 305
Db 243 KATICEANMPTLVDOGFWMIGGOYALPTSLPODYVSEAYALHTLFARPPSSAALAFVNA 302
Qy 306 GLPQGTAPAGTAPAEQASGGYLTWRHNGTTFPAGSVSYLPEGFALERYDND 365
Db 303 SMPGGSAPGSDPAMIPDSTGFC-QWRHGFADPAGVITVITPRGYTMOVFTTNNEMNG 361
Qy 366 FASAGDTVFRQVAADVEVYVNNPAGG-----SAPFTVRYV-PPSNATYTVRNTL 418
Db 362 FANPDVYTFGQ-----TGGAAGTMTITTAFTVTLTATTSAAVINFENLDA 413
Qy 419 ETRPSSRRLELPMPPADFGQTVANNPKIEOSLKEITLGCYLVHSMKRNPFOLTPASSFG 478
Db 414 ETTAASNRSEVPLPLPFGOTAPANNPKIEOTLVKDTLGCYLVHSMKRNPFOLTPASSFG 473
Qy 479 AVSFNNPGYERTRLPDYTGIRDSFDQNMSTAVAHFRSLSHSCSIYTKYQGWEGVTNVP 538
Db 474 AISFTNPGFGRNLDLPFGFGRISLDVNMSTAVCHFRSLSHSCSIYTKYQGWEGVTNVP 533
Qy 539 TPRGOFAGHGLKNEIELCLADDLATRLTGYVPATDNFAAASAFANMLSSVLKSEATS 598
Db 534 TPRGOFAGHGLKNEIELCLADDLATRLTGYVPATDNFAAASAFANMLSSVLKSEATS 593
Qy 599 SIKSVGETAVGAOSGLAKLPGLIMSVPGKTAARVARRARRARRARAN 647
Db 594 SVIKELGNQATGLANQGLARLPGLASIPGKTAARVARRARRARRARAN 642
```

```
RESULT 3
Q9YK87 PRELIMINARY: PRT: 757 AA.
ID 09YK87
AC 09YK87;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CAPSID PROTEIN.
GN TAV-CP.
OS Thosea asigna virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
OC unclassified Tetraviridae.
OX NCB1_Taxid=83810;
RN RN
RP SEQUENCE FROM N.A.
RX MEDLINE=99350019; PubMed=10423156;
RA Pringle F.M., Gordon K.H., Hanzlik T.N., Kalmakoff J., Scotti P.D.,
Ward V.K.;
RT "A novel capsid expression strategy for Thosea asigna virus
(Tetraviridae).";
RL J. Gen. Virol. 80:1855-1863(1999).
DR EMBL; AF062037; AAC97195.1; -.
DR MEROPS: A21.UNM; -.
FT CHAIN 1 155 UNKNOWN.
FT CHAIN 156 695 LARGE CAPSID PROTEIN.
FT CHAIN 696 757 SMALL CAPSID PROTEIN.
SQ SEQUENCE 757 AA: 82413 MW: 17432C70D2C427C CRC64;
```

Query Match 12.9%; Score 435.5; DB 12; Length 757;  
Best Local Similarity 25.9%; Pred. No. 2.4e-22;  
Matches 170; Conservative 105; Mismatches 263; Indels 119; Gaps 27;

```
Qy 26 TYVNGRRNORR---TGROVSPDNFTAAADLQSLDANTYTFPANISMPDEFNRMAK 82
Db 174 TKTAPGKTRRRRRNRCKGKIGNPNPFSSISA-----GAPMSKIPALRGTLA 219
```

```
OY 83 GKIDSDSIGWYFKYLDPAGATESARAVGEYSKIPDGLVKSVDAREIENECPVYT 142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 MREPISSPVSGVLETLNLDPCGEYKSTL---DYKVPDGAIPMSVCGQFRETITIRHPCAN 276
OY 143 DVSVPDGRGWSLSIFSEPMFRTAVAVANENKEMSLDYNDLIEMLNN---LADWRY 198
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 PALAPDGSMTLGIIRTLNMTPLLLADMNNAIESDSELDVARSFNNYGLADL--- 333
OY 199 VVDSQWQINFTNDTTYRIRIRLRPTDYDPDE-GLVRTVSDYILTYKATICEANMPL 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 -ASHPNWVE-TSQAGLYNSIQHWITALIGPPSDTGVSPDYITDFRITGSGFVSSNTBPL 391
OY 258 VDOGFWMIGQVYALTPTS-----LPQDYSEAVALTITFAPRSSAAALAFVAGLPOGT 312
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 392 INQGNVVAQF--NPNTKEKVIPIHSESEETPLGWSRLKSVASVAVIITYIOPGV--GGA 447
OY 313 APAGTPAMEQASSGGYLTWRHNGTTFPAGSV--SYVLPEGFALERYDPNDGSMWTFASAG 370
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 448 QNQRIV-----TLNRSGSVINSSVLPO--VTQGYRFENNG---DTYAVG 485
OY 371 DVTTRQYAVADVVTYNNPAGGSAPFTVRYRPPS--NAYTN-----TVFRN----- 415
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 DTLRFQ-----LTERDSGG---YTVRLQRSADGTFYINIGLPLSTTGTGONGGFNTD 532
OY 416 ----TLLETTPSSRLLE-LMPMPADFGQTVANNPKIEOSLKETLGCYLVHSMKNRPYQ 470
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 533 SGLGVLLEGDVLAGRVNMLTLPTFQSDLMQOTPKTCVFOIKE--GGFYVRDDIQOPYN 590
OY 471 LTPASSFGAVSFNNPDEYERTD-LPDYTG-IRDSFDQNMSTAVAHFRSLSHCSIVTKY 528
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 591 MTPASRYAPVREVSNAI--TLDALNSAVGTIRDTADSNYGALCHMTSLPLACAFIKAE 648
OY 529 QGMEVTVNNPFGQFAHAGLLKNEELICLADDLATRLGVYPAIDNFPAASAPRANML 588
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 649 MREFEAVPGRNSPWFPAJSTPPKDEVALTISTVMDLEPFPAPLAYNGMLFSVLSIV 708
OY 589 SSVLSEATSSIIKSVGETAVGA---AOSGLALPGLMSVYGKIAARYARRARRAR 642
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 709 RRIPALRTRAS--NVAETVAGCVEDAQAGVT-----KYETVYRGRGAMR 751
```

RESULT 4

O9YRB2 PRELIMINARY: PRT: 612 AA.

AC O9YRB2: 01-MAY-1999 (Tremblrel. 10, Created)

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE CAPSID PROTEIN.

OS Nudarelia capensis beta virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae; Betateiravirus.

OC NCB1\_Taxid=85652;

OX NCB1\_Taxid=85652;

RN [1]

RN SEQUENCE FROM N.A.

RX MEDLINE=99263183; PubMed=10329566;

RA Gordon K.H., Williams M.R., Hendry D.A., Hanzlik T.N.:

RT "Sequence of the genomic RNA of nudarelia beta virus (Tetraviridae) defines a novel virus genome organization."

RL Virology 258:42-53(1999).

RN [2]

RN SEQUENCE FROM N.A.

RP Gordon K.H.J., Williams M.R., Hendry D.A., Hanzlik T.N.:

RA Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF102884; AAC97510.1; --

DR MEROPS: A21.0MW; --

SO SEQUENCE 612 AA; 66406 MW; 0E34029D3B39085 CRC64;

Query Match 11.5%; Score 386.5; DB 12; Length 612;  
Best Local Similarity 26.0%; Pred No. 5.2e-19;  
Matches 174; Conservative 86; Mismatches 295; Indels 115; Gaps 31;

```
OY 15 RCTRNVASANTVTYNGRRNORRPTGRQVSP---PDNFTAAQDLAQSLLDANTYFPANI 71
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 RPARNNPSOGNQGRRNNNNRRRRRGLKLPVAD--ITSQGMAPNNHNT----- 58
OY 72 SSMPEFRWAKG-----KIDSDSIGWYFKYLDPAGATESARAVGEYSKIPDGLVKS 125
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 -RVNNGRRVRRLQAMMESPAAATSEAMIHDIYLDPDDEYKTSLDG---KIPDAIQOS 114
OY 126 VDAETREIYNECPVYTVSDPLDGRGWSLSIFSEPMFRTAVAVANENKEMSLDYN 184
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 TCGQFRGVGARYPALNSTLPLDGTWPLVMHLPFRRHPLLFTTTSNTE--VEYINA 172
OY 185 DL-----TEMLNLAAMRYVVOSEQWINTDNTTYVYRIVLRPT---YDVDPPE-GLVR 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 DIDAPANDH-NRRTWTEAT-YPSAQGN--VEFYMV-----PTEALDPPPTQLGVSG 224
OY 237 TVSDYRLTYKATICEANPFLVDGFWIGQYALTPSLPODYSEAVALTLT----- 290
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 LLESYRLTSSGVYAVFNAPTLVNOGVAVIAQF--QPKHEQKENDDIVAGTQTGTQL 282
OY 291 -PARSSAAALAFVAGLPQGTG-PAGTPAMEQASSGGYLTWRHNGTTFPAG---SVSY 345
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 GSGGPNYTLTWT-IDDOVEFGAALPLPTVSMGPMBESGQIVFYQANLTFDVGNTITTT 341
OY 346 VLPEGFALERYDPNDGSMWTFASAG-DVTFRQYAVADVVTYNNPAGGSAPFTVRYRPP 404
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 342 TLPPGSV-----TGMOFTASNGTDVY-----VD-----ACARLYAFANIDA 380
OY 405 SNAVYNTVFRNLTLETTPSSRLLELMPMPADFGQTVANNPKIEOSLKETLGCYLVHSM 464
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 381 SELNLQDI-----NSIKIPPTNMNQWQATPKIQQLNETKGFTV-PLRA 425
OY 465 RNPVQQLPAPSSFGAVSFNNPDEYERTDLPY-----TGIRDSFDQNMSTAVAHFSLHS 520
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 FQVPEMTMATSYGVRMKTP--RT-TYVDHRAIGLQIDIDSNFAIGVAAAMTGMSTS 481
OY 521 CSIVTKTYQGMEVTVNNPFGQFAHAGLLKNEELICLADDLATRLGVYPAIDNFPAAY 580
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 482 TVPYKRVRRFPAIPABESPMPFPASATPPKDDVALTYARWTDLHPAYREYRGFPAL 541
OY 581 SAFAANMLSSVLSKSEATSSIIKSVGETAVGAOSGLAKLPGLMSVPGKIAARYARRAR 640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 542 FAMVAKTIAQLPR-----YVRS---AAGVANAVTDCIESATESVANSSTERRORAR 591
OY 641 R-----RAAR 645
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 592 RVGIGARGAR 601
```

RESULT 5

O9L950 PRELIMINARY: PRT: 2147 AA.

AC O9L950: 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE OTHER MEMBRANE-LIKE PROTEIN.

OS Pseudomonas putida.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.

OC NCB1\_Taxid=303;

OX NCB1\_Taxid=303;

RN [1]

RN SEQUENCE FROM N.A.

RX MEDLINE=20225830; PubMed=10762233;

RA Espinosa-Urgel M., Salido A., Ramos J.L.:

RT "Genetic analysis of functions involved in adhesion of pseudomonas putida to seeds."

RL J. Bacteriol. 182:2363-2369(2000).

RN [2]

RN SEQUENCE FROM N.A.

RP Espinosa-Urgel M., Salido A., Ramos J.L.:

RA Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF182515; AAF69025.1; -  
 DR InterPro: IPR001343; HemIysn\_Ca\_bind.  
 DR PRINTS: PRO013; CABNDNGRPT.  
 SQ SEQUENCE 2147 AA; 211587 MW; BD7EC3B460FE2D6B CRC64;

Query Match 4.08; Score 134; DB 2; Length 2147;  
 Best Local Similarity 22.08; Pred. No. 2.8;  
 Matches 141; Conservative 75; Mismatches 264; Indels 162; Gaps 33;

QY 39 TGRVSPDP-NFTAAODLAQSLDANTVTEPPANISSMPEFRNMAKGRIDSDSI-----GW 94  
 DB 1363 TGRVLPDLTPPAAGLFLFSADGGLTGTGEAVATVNS-AGGV-LQTATVQADGT 1420  
 QY 95 YFKYLDAGATASARAVGESKIPDGLVKSVDAREIREINEECPPVTD-----VSPLD 149  
 DB 1421 FTVTLSP--PDONGILLSVTLSDPRGNVSGVNVYATPDV-DANPVIASDLATATVNL 1477  
 QY 150 G-----RWSLSTSFPMFRFAVYAVANVENKESLDVYNDLIEMNLADMRVVD 201  
 DB 1478 AVSATKTYTDSFTLLSGFSKTYMTVAAGTTADPTLTL-TTNSVLAALLNTSFTLQVKD 1536  
 QY 202 -SEQMINFTNDTYYVRLRVLPDYDPDTEGL-----VTVSDYRLTYKA--ITCEAN 253  
 DB 1537 ASGANVITATGNTGGLDLVLP-----GLQVDIGVLQAGDIRLVGSGGIGLITE 1588  
 QY 254 MPTLVDOGFWIGQYALTPSLPY-----DVSEAYALHTLTFAR-PSSAA 299  
 DB 1589 VSTLID-----IVSTSLTQFTGTGINTSGNVITDVTGCAVD--ARCPDSAAV 1634  
 QY 300 LAFWAGLIPQGGTAPACTPAMEQASSG-----GYLWRKNGTTFAGSV--SYL-- 347  
 DB 1635 LQV-----LKDSYVSAAGTATTVQOQYGLVIRADGSYTYTPNGSPNSGVKDYFSLVH 1690  
 QY 348 PEGR--ALERY-----DPNDGSWTDFASAGDYTFTEQVAVD-EVVYTNPNAG 391  
 DB 1691 PNGISDANLYVRIDSPQATFVSDTYNGAPATVYADVNDVGTQVYLLDNKRVDSSTLG 1750  
 QY 392 GGSAPFTTVRPPSNATVTFKNTLTLEPRSSRLLEPMRPADGGOTVANNPKIEOSL 451  
 DB 1751 SLNLPVIGTR--QATYTTTVAANTTADLQVNVSTNL-----LSLNGTTEIELKL 1799  
 QY 452 KETLGCYLVHSKMRNPFOULPRASSFGAVSFNNPCYERTNLPYTLIRSFQDNMTAV 511  
 DB 1800 NPATGQTV-----LVQSVPGSLVSLGGAGY-----TF-ENQAGT 1836  
 QY 512 AHFSLSHSCSIVTKYQEGEVTVNTP--FGQFAHAGLKNELICLADLADLRLTG 568  
 DB 1837 YHNVATAGGIGLSSI-----TTSLNITTTTYLLEFYVSG-----ATAATG 1876  
 QY 569 VVPATDNFAAASAFANMLSSVLKSEATSSIIKSVGETAVG 610  
 DB 1877 NLADDMVGSALTIVL-----SVLTAAVTYTIIPYNGSVSAG 1912

RESULT 6  
 Q90P87 PRELIMINARY; PRT; 886 AA.

AC Q90P87;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE MAJOR OUTER ENVELOPE GLYCOPROTEIN GP350.  
 OS Human herpesvirus 4 (Epstein-Barr virus).  
 OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 NCBI\_TaxID=10376;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98203772; PubMed=9542645;  
 RA Lee W.K., Kim S.M., Sim Y.S., Cho S.G., Park S.H., Kim C.W.,  
 RA Park J.G.;  
 RT "B-lymphoblastoid cell lines from cancer patients.";

RL In vitro Cell. Dev. Biol. Anim. 34:97-100(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99072166; PubMed=9856346;  
 RA Chang S.H., Kim S.H., Lee W.K., Kim H.J., Choi S.H., Park J.H.,  
 RA Jang H.S., Chung G.H., Kwon T.H., Kim D.H., Yang M.S., Jang Y.S.;  
 RT Cloning and analysis of the Epstein-Barr virus glycoprotein 350  
 RT genes.";  
 RL Mol. Cells 8:585-593(1998).

RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Chang S.H., Kim S.H., Lee W.K., Kim H.J., Choi S.H., Park J.H.,  
 RA Jang H.S., Chung G.H., Kwon T.H., Kim D.H., Yang M.S., Jang Y.S.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF172332; AAD51697.1; -  
 DR InterPro: IPR002965; P-rich\_extensn.  
 DR PRINTS: PRO1217; PRICHEXTENSU.  
 KW Envelope protein.  
 SQ SEQUENCE 886 AA; 92487 MW; F4BCA573EBE29A48 CRC64;

Query Match 3.9%; Score 132; DB 12; Length 886;  
 Best Local Similarity 20.5%; Pred. No. 1;  
 Matches 139; Conservative 66; Mismatches 268; Indels 206; Gaps 29;

QY 48 NFTAADLAQSLDANTVTEPPANISSMPEFRNMAKGRIDSDSIGWYFYLDPAQATES 107  
 DB 169 NITAVNR-AQGLD--VTLPLSLTPSAQDSNFS-----VKTEMG--NEID---IEC 211  
 QY 108 ARAVGESKIPDGLVKSVDARE-----IYNECPVTVDSVPLDGRKSLSTFSF 160  
 DB 212 IMEDGELISQVLPQGNKNITVCSGESHVPSGGILTSTSPAT--PLPGTAYASLRLTPR 269  
 QY 161 PMFR-----TAYVAVANVENKESLD-----VVNDLI-----EWNLNADMRVYDSE 203  
 DB 270 PVSRLFNNSILYFYSGNGPKASGDYCIQSNIIVFEDELIPASODMPTNTDITLYGDN- 328  
 QY 204 QWINTFTDTYYVIRVRLRPTVDPDTEGLVTVSDYRLTYAITCEANMPLVDGFW 263  
 DB 329 -----ATYSVPMTSE-----DANSPVMTVAFW 352  
 QY 264 -----ICGOYALPPT----- 273  
 DB 353 AMPNNETDKCKWTLSGTPSCGENTSGAFASNRFTDITVSGLGTAPKTLITTRATNA 412  
 QY 274 -----SLPODYSEAYALHTLTFARPSAALAFVWAGLPQGGTAPAGTPAMEQAS 324  
 DB 413 TTTTHKVIYFSKAPESSTTSPLMTTGFAAPNT-----TGLPSSHVPTNLTA--PIS 463  
 QY 325 SGGYLTWRHNGTTPPAGSVYLPBGFALERYPDNG-----SWDFASAGDYTFRQ 377  
 DB 464 TGPVSTADYTSPTPAQTSGASP--VMPRPSPRNGTESKAPDMTSPSAVTTTPPNA 520  
 QY 378 VAVDEVVVT-----NNPAGGSAPTFTVRVPNNATVNT-----VERNLTLEFRPS 423  
 DB 521 TSETPAVTTPTPNATSPFLTKTSPTSATVTPPTPNATIPTPNATIPPLGKITSF 560  
 QY 424 SRRLELPMPPADGGOTVANNPKIEOSLKEKTCYLVHSKMRNPVQLTPASSFGAVSFN 483  
 DB 581 S-AVTPPTPNATSPYGETSPKANTT--NHITLG-----GTSSTPVYTSPPKNTSAVTTG 632  
 QY 484 NPGYERTDLPDYTGIR-DSFDONMSTAVAHFPSLSHSCSIVTKYQEGEVTVNTPPG 542  
 DB 633 Q--HNITSSSTSMISLRPSISSETLSPNSD-NSMSHMLPLTSAHPTGGENITQV--TPAS 688  
 QY 543 QFAHAGLKNELICLADLADLRLTGVPAT-----DNFAAASAFANMLSSVLKSE 555  
 DB 689 TSTH-----HVTSSPAPRPDTISQASGPNSSSTKRGVEVNTKGTTPPKN 734  
 QY 596 ATSSIIKSVGETAVGAOS 614  
 DB 735 ATSPQAPSGOKTAVPTVTS 753

```

RESULT 7
ID 056854 PRELIMINARY: PRT: 856 AA.
AC 056854:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GP340.
GN BLRF1.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
RX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M81;
RX MEDLINE=97082049; PubMed=8923292;
RA Mackelt M., Cox C., Pepper S.D.V., Lees J.F., Beverley B.A.,
RA Wedderburn N., Arrand J.R.;
RT "Immunization of common marmosets with vaccinia virus expressing
RT Epstein-Barr virus (EBV) gp340 and challenge with EBV.";
RL J. Med. Virol. 50:263-271(1996).
DR EMBL; X99106; CA67558.1;
SQ SEQUENCE 856 AA; 89634 MW; E34E581AF1071CDD CRC64;

```

```

Query Match 3.9%; Score 131.5; DB 12; Length 856;
Best Local Similarity 20.8%; Pred. No. 1;
Matches 120; Conservative 51; Mismatches 190; Indels 217; Gaps 23;

```

```

QY 48 NPTAAADLAQSLDANTVTFPANISSMPEFRMA-----KKRIDSDSIGMYEKYLD 101
DB 169 NITAVVR--AGLD--VTLPLSLPTSADSNFSVKTOMLGNLEID----- 209
QY 102 ACATESARAVGEYSKIPDGLVFSVDAIRE-----IYNECPYVDVSVPLDGRQWS 154
DB 210 ----ECIMEDGEISOVLPDKNFNITSCGESHVPSGGILSTSPVAT--PIPGGAYAS 263
QY 155 LSFSPFMR-----TAVVAVANENKMSLD-----VYNDLI-----EMLNLMADR 197
DB 264 LILTRPVSRFGLNNSILYFSGNPKRSGDYCIQSNIYSDELPAQDMPTNTTDT 323
QY 198 YVVDSEQWINTNDNTYVYRIVLRPTVDVDPTEGLVTVSDYRLTYKAITCEANMPTL 257
DB 324 YVGDN-----ATYSVPWVTSF-----DANSPPV 346
QY 258 VVOGW-----IGGOYALTP----- 273
DB 347 TVTAFWAMENNTETDEKCKMTLTSCTPSCENISGAFASNRFTDITVSGLTAPKTLIT 406
QY 274 -----SLPODYSEAYALHTLFARPSAAALAFWAGLPGGTAPAGT- 317
DB 407 KRAITNAITTHKVIKSKAPESTTTSPTLTWTFGADBNITGTPSS-THVPTMLTAPASTG 465
QY 318 -----PAMEQASSGGYLW-----RHNGT-----TFPAGSVSYVLEPEG- 351
DB 466 PTVSTADVTSPAGTTSASPVTPSPRDNGTESKAPDMSPISAVTTPPNNGSPRP 525
QY 352 ALERIDPNDG-----WTPDASAGDVTFRQYAVDEVYVYTNPDAG----- 391
DB 526 AMTTPPNATSPGLKTSPTSAVTTPTPNATSPTPAVTTPPNATSPITVGETSPQANATN 585
QY 392 ---GSAFTYRVPPSNAYTNTVFRNTLLETRPSSRRLLEMPADFCQYVANNPKLQ 448
DB 586 HILGGSFPPVYTPPKNATSDV---TTGQHNRTSSSTSSKLRPSSIPETTSHPMLTS 642
QY 449 SILKETLGCYLVHSKMRNPVQLTPAS-SFGAVSFNNP 485
DB 643 A-----HPTGGENITGVTPASISTHNVSTSP 669

```

RESULT 8

```

Q9RY11
ID 09RY11 PRELIMINARY: PRT: 597 AA.
AC 09RY11:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE HYPOTHETICAL 60.6 KDA PROTEIN.
GN DR0143.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.U., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001876; AAF09732.1;
DR TIGR; DR0143;
DR InterPro; IPR001917; AminoTransf_2.
DR InterPro; IPR001344; DUF11.
DR Pfam; PF01345; DUF11.
DR ProSite; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 597 AA; 60604 MW; B9A61442CE8A8279 CRC64;

```

```

Query Match 3.9%; Score 130; DB 16; Length 597;
Best Local Similarity 20.6%; Pred. No. 0.76;
Matches 128; Conservative 67; Mismatches 253; Indels 172; Gaps 25;

```

```

QY 54 ODLAQSLDANTVTFPANISSMPEFR-----NMAKKRIDSDSIGMYEKYLDPAAGT----- 105
DB 4 RDIYDLPAPAGGGGVVSTSPYQAVCVNNSGGILSNSSPAVSLTAPLTDNIG--NGLETGVTT 163
QY 106 ----ESARAVGEYSKIPDGLVFSVDAIRE--IYNECP--VYVDSEVPLDGRQWSLSI 157
DB 64 PTAGATQMSVPNTIIPACTCRFSFDVTLPRAGTYKNYIPAIRRSGGNDAGNASTTA 123
QY 158 FSPFMRITAY-----VAVANENKMSLDV--VNDLIMLNLMADRYV 200
DB 124 FAGPTVTKAYSPKSLDGKATSTLITLNNSSPAVSLTAPLTDNIG--NGLETGVTT 181
QY 201 DSEQWINTNDNTYVYRIVLRP-----YVDPDTEGLVTVSDYRLTYKAITCEANMP 255
DB 182 SCPTATPSTGTTIYTPSGATLNPCTTATVRSATAG-----SY 222
QY 256 TLVDG---FWIGGOYA-----LTPSLPODYSEAYA-----LHTLFARPSA 297
DB 223 NRTSAGALQTYGNNAAASDLTYTSTRLITTHASQNTAGOTGYTYVSNASGA 282
QY 298 AALAFVWAGLPGGTAPAGTAPAMEQASSGGYLWHRNGTTFPAGSVSYVLEPEGALERD 357
DB 283 AA-----TSGALSLDLSGMSFNSVT 305
QY 358 PN-DGSMWTFDASAGDVTFRQYAVDEVYVYTNPDAGGSAFTTVVPPSNATNTVFNKT 416
DB 306 TTAGGSFGTRPSAGAT-----GRVDWTFPTSPPLAAGSLTFETVVVANNVANNATLTN 360
QY 417 LLETSPSSRRLLEMPPAD-FGQYVANNPKLEOSILKETLGCYLVHSKMRNPVQLT--- 472
DB 361 VASVGGGDPDVLTPFGATCTGEOCASDPTTY-----KRITQTLIRK 402
QY 473 --PASSFGAVSFNNPGEYERTRDLPDYGI RDSFDQNMST-AVAHFRSLSHSCSIVTKTYQ 529

```

Db 403 EFGGAGPKGEGN-----YDVATI--TISGSSLTLTATSSIANCPGVAITDT-- 449  
QY 530 GWEGVTWN-TPEGQFAGHAGLKNKEIICLADLATRLTGYPATDNFAAFAANM- 587  
Db 450 -----INIRPGANYTLRELVKNDVAFSGPDSYDSR-----YTCTNMTGTSTTWPINSS 498  
QY 588 -LSSVLKSEATSSIIKSVE 606  
Db 499 GMSFTLTPOAGDIITCSVD 518  
RESULT 9  
ID Q95S10 PRELIMINARY; PRT; 466 AA.  
AC Q95S10;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE H10238P.  
-GN DSCAM.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
EX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y, CN BW SP;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Chame M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Pacled J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY061009; AAL28557.1; -  
SQ SEQUENCE 466 AA; 50943 MW; 7056AB89B955CC2B CRC64;  
Query Match 3.8%; Score 127; DB 5; Length 466;  
Best Local Similarity 20.7%; Pred. No. 0.84;  
Matches 77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;  
QY 231 TEGIVRTVSRYRLTYKATICEANMPTLVDOGFMT-----GGQVALTPSLPQYDVSEA 283  
Db 71 TETKTQVPHYQMSFEATELEKNRP---YEFWTASTTIGEGQSKSIYAMPDDOYPAK 126  
QY 284 YA-----LHTLFARPPSSAALAFVWAGLPDGGTAPAGTAPMEQASSGGYLTWRHNGTTP 339  
Db 127 IASFDFTFTAFKEDAKMPLA---VGAPQ-----PEITWIKGVERS 166  
QY 340 AGSVSYLPRGFL-----ERYDPNDS-WTDFASAGDTYTFROYAV----- 380  
Db 167 ANDRMRYLPDGSLLIKSVNRQDAGDYSCHAENSIAKDSITHKLIIVLAPQSPHVTLSATT 226  
QY 381 -DEVVYTNPNAGGSAPT--FTVRVPPS-----NAYTN 410  
Db 227 TDALITVAKLKHEDTPALHGYTLHYKPEFGEWETSEVSDQKNIEGLICSGRYOYA- 285  
QY 411 TVERN-----TLETRPSSRLLEMPRADGQTIVANNPKIEOSLKEITLGC---YL 459  
Db 286 TGFNNTIGAGBASDILNRTRGOKPLPEKPR-FIEVSSNSVSLHFKWMDG-CGPMSHV 343  
QY 460 VHSKMRNP-----VFQLTPASSFG--AVSFNNPGY----- 487  
Db 344 VESKKRDOIEWNOISNNVKNPNVYVLDLEPATWYMLRITAHNSAGFTVAEYDFATLTVT 403  
QY 488 -----ERTRLPD 495  
Db 404 GGTIAPSRLPE 415  
RESULT 10  
Q95TGO

ID Q95TGO PRELIMINARY; PRT; 1212 AA.  
AC Q95TGO;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE S09407P.  
-GN DSCAM.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
EX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y, CN BW SP;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Chame M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Pacled J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY058789; AAL14018.1; -  
SQ SEQUENCE 1212 AA; 134066 MW; 6D9061E62EF9BB5C CRC64;  
Query Match 3.8%; Score 127; DB 5; Length 1212;  
Best Local Similarity 20.7%; Pred. No. 3.6;  
Matches 77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;  
QY 231 TEGIVRTVSRYRLTYKATICEANMPTLVDOGFMT-----GGQVALTPSLPQYDVSEA 283  
Db 455 TETKTQVPHYQMSFEATELEKNRP---YEFWTASTTIGEGQSKSIYAMPDDOYPAK 510  
QY 284 YA-----LHTLFARPPSSAALAFVWAGLPDGGTAPAGTAPMEQASSGGYLTWRHNGTTP 339  
Db 511 IASFDFTFTAFKEDAKMPLA---VGAPQ-----PEITWIKGVERS 550  
QY 340 AGSVSYLPRGFL-----ERYDPNDS-WTDFASAGDTYTFROYAV----- 380  
Db 551 ANDRMRYLPDGSLLIKSVNRQDAGDYSCHAENSIAKDSITHKLIIVLAPQSPHVTLSATT 610  
QY 381 -DEVVYTNPNAGGSAPT--FTVRVPPS-----NAYTN 410  
Db 611 TDALITVAKLKHEDTPALHGYTLHYKPEFGEWETSEVSDQKNIEGLICSGRYOYA- 669  
QY 411 TVERN-----TLETRPSSRLLEMPRADGQTIVANNPKIEOSLKEITLGC---YL 459  
Db 670 TGFNNTIGAGBASDILNRTRGOKPLPEKPR-FIEVSSNSVSLHFKWMDG-CGPMSHV 727  
QY 460 VHSKMRNP-----VFQLTPASSFG--AVSFNNPGY----- 487  
Db 728 VESKKRDOIEWNOISNNVKNPNVYVLDLEPATWYMLRITAHNSAGFTVAEYDFATLTVT 787  
QY 488 -----ERTRLPD 495  
Db 788 GGTIAPSRLPE 799  
RESULT 11  
Q9VAJ9  
ID Q9VAJ9 PRELIMINARY; PRT; 2016 AA.  
AC Q9VAJ9;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CG17800 PROTEIN.  
-GN DSCAM OR CG17800.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
EX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed-10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amarantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borova D., Botchan M.A., Bouck J., Brokstein P., Brotlier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclik J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003841; AAF59271.1;  
DR HSSP: P40189; 180U.  
DR FlyBase: FBgn0033159; Dscam.  
DR InterPro: IPR000267; Asparagine\_glutamase.  
DR InterPro: IPR003961; FN\_III.  
DR InterPro: IPR003598; IG\_c2.  
DR InterPro: IPR003600; IG\_1like.  
DR InterPro: IPR003006; IG\_MHC.  
DR Pfam: PF00041; fn3; 6.  
DR Pfam: PF00047; Ig; 10.  
DR SMART: SM00060; FN3; 6.  
DR SMART: SM00408; IGC2; 7.  
DR SMART: SM00410; IG\_1like; 2.  
DR PROSITE: PS00444; ASN\_GLN\_ASE\_1; UNKNOWN\_1.  
KW Immunoglobulin domain.  
SQ SEQUENCE 2016 AA; 222109 MW; 64A8DE3B7BD0AB0 CRC64;

Query Match 3.8%; Score 127; DB 5; Length 2016;  
Best Local Similarity 20.7%; Pred. No. 7.9;  
Matches 77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;

OY 231 TEGLVRTVSDRLTKATTCCEANMPTLVDOGFMI-----GGQYALPTSLPQYDVSEA 283  
DB 1259 TETKTQKVPHYOMSEATELEKNRP-----YEFWVASTTIGGGQSKSVAMPSPQYPAK 1314  
OY 284 YA-----LHTLTFAFPSSAAALAFWAGLPQGGTAPAGTPAWEQASSGGYLTWRHNGTTFP 339  
DB 1315 IASFDFTFAATFKEDAKMPCLA---VGAPQ-----PEITWKIKGVESF 1354  
OY 340 AGSVSYVLPGEFAL-----ERYDPNDGS-WTDFASAGDTVTFRQYAV----- 380  
DB 1355 ANDRRKRVLPDGSLLIKSVNRQDAGDYSCHAENSIKADSTTHKLIVLAPPOSPHTLSATTT 1414

OY 381 -DEVVTVNNPAGGSAPT--FTVRVPSS-----NAYTN 410  
DB 1415 TDLTVTKAKPHEGDTAPLHGTLTKPERGEWSESVSDOKNIEBLLGSRQYVA 1473  
OY 411 TVFRN-----TLLETRPSSRRLELPMPPADGQTVANNPKFIESLKETLGC----YL 459  
DB 1474 TGFNNIGAGEASDILNTRKKGQKPLPEKPR-FIEVSSNSVSLHKRKKDG-GCPMSHFV 1531  
OY 460 VSKSRNP-----VFQLTPASSFG--AVSFNNPGY----- 487  
DB 1532 VSKSRKRDQEMWQISNNVKNPDNNYVLLDEPATWYNLRTAHNSAGFTVAEYFATLTVT 1591  
OY 488 -----ERTRDLPD 495  
DB 1592 GGTIAPSRDLPE 1603

RESULT 12  
G9NBAL  
ID G9NBAL PRELIMINARY; PRT: 2016 AA.  
AC G9NBAL: 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE DSCAM PRECURSOR.  
GN DSCAM OR CG17800.  
OS *Drosophila melanogaster* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; *Drosophilidae*; *Drosophila*.  
NCBI\_TaxID=7227;  
RP [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=20348742; PubMed-10892653;  
RA Schmucker D., Clemens J.C., Shu H., Worby C.A., Xiao J., Muda M.,  
RA Dixon J.E., Zipursky S.L.;  
RT "Drosophila Dscam is an Axon Guidance Receptor Exhibiting  
RT Extraordinary Molecular Diversity.";  
RL Cell 101:671-684(2000).  
DR EMBL: AF260530; AAF71926.1;  
DR HSSP: P40189; 180U.  
DR FlyBase: FBgn0033159; Dscam.  
DR InterPro: IPR000267; Asparagine\_glutamase.  
DR InterPro: IPR003961; FN\_III.  
DR InterPro: IPR003598; IG\_c2.  
DR InterPro: IPR003600; IG\_1like.  
DR InterPro: IPR003006; IG\_MHC.  
DR Pfam: PF00041; fn3; 6.  
DR Pfam: PF00047; Ig; 10.  
DR SMART: SM00060; FN3; 6.  
DR SMART: SM00408; IGC2; 7.  
DR SMART: SM00410; IG\_1like; 3.  
DR PROSITE: PS00444; ASN\_GLN\_ASE\_1; UNKNOWN\_1.  
KW Immunoglobulin domain; Signal.  
FT SIGNAL 1  
SQ SEQUENCE 2016 AA; 222124 MW; 95CF95488F2AD36C CRC64;

Query Match 3.8%; Score 127; DB 5; Length 2016;  
Best Local Similarity 20.7%; Pred. No. 7.9;  
Matches 77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;

OY 231 TEGLVRTVSDRLTKATTCCEANMPTLVDOGFMI-----GGQYALPTSLPQYDVSEA 283  
DB 1259 TETKTQKVPHYOMSEATELEKNRP-----YEFWVASTTIGGGQSKSVAMPSPQYPAK 1314  
OY 284 YA-----LHTLTFAFPSSAAALAFWAGLPQGGTAPAGTPAWEQASSGGYLTWRHNGTTFP 339  
DB 1315 IASFDFTFAATFKEDAKMPCLA---VGAPQ-----PEITWKIKGVESF 1354  
OY 340 AGSVSYVLPGEFAL-----ERYDPNDGS-WTDFASAGDTVTFRQYAV----- 380





```

QY 413 FRNTL-LETRPSSRLLELPPADFGQFVANNPKIEOSLKEETL-----GCYLVRK 463
DB 477 TASSVTIAMNASTDNVGT-----GYTYNGTSLVSTGTATITISGLAPSTTFYTK 530
QY 464 MNPVQQLTPASSFGAVSFNNPGYERTDLPDYGIKRSFDQNMSTAVAHFSLSHSCSI 523
DB 531 AKDAGNLSAASNSLTVS-----TTVQPGGDTQAPTPVPTNLSTFAKTSST 575
QY 524 VKKT-----YQMGVTVNTPFGQFAH-AGLKNKEELLCLADLATRLTGIV 570
DB 576 IFLSAASTDNVGVIGVEYNGTALVTYVSGTSATVTGL-----TADTSYTFYVAK 627
QY 571 PATDNFAAFAANMNLSEVLSKSEATSSIIKSGET--AVGAQSGLAKLPLMSVPG 628
DB 628 DAAGNLASAASSAL-----TYKTE-----VGTTPNGVSAMKANATAYVQGLVTVNG 672
QY 629 K 629
DB 673 K 673

```

## RESULT 15

```

Q97206 PRELIMINARY; PRT; 1269 AA.
AC Q97206;
DT 01-OCT-2001 (TRENBLREL. 18, Created)
DT 01-OCT-2001 (TRENBLREL. 18, Last sequence update)
DT 01-OCT-2001 (TRENBLREL. 18, Last annotation update)
DE PROTEASE RELATED PROTEIN.
GN SS01141.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyer M.J., Chan-Weiher C.C.Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder F., Schenk M.E., Theriault C., Tolstrup N.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL: AE006730; AAK1392.1;
DR InterPro: IPR000209; Peptidase_S8.
DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1269 AA; 138027 MW; 87C8F05CFCF83DA84 CRC64;

```

Query Match 3.7%, Score 125.5; DB 17; Length 1269;  
 Best Local Similarity 19.6%, Pred. NO. 5;  
 Matches 98; Conservative 69; Mismatches 181; Indels 153; Gaps 27;

```

QY 18 RNVRSATFVTVNGRRNRRTGRGVSPDNTFAAODLAOSLDANTVTFPANISSMPEF 77
DB 584 QRLVTAITITPNG-----SPVQ--TGEFFALIEYVGLNLTFTNTYLSLTKL 629
QY 78 RNMAGKIDLDSDSIGWFKYLDPAGATESARAVEYSKIPDGLVFEVDAREIRY--- 134
DB 630 --WT-GSGVLNKAAGLFFVYV--YGSSDGLRGIGYTFEFGYITFTFYVVEL 684
QY 135 -NEECPVTVDS---VPLDGQWMSLIFSFPMEFRTAIVAVAN---VENKESLDVNDL 186
DB 685 GNAELGITLSNSYFQAPIGVNNITLNTSYNTTNAFYFTLLSVKNGVAVIDLPPDL 744
QY 187 IEWLNLADWRYVVDSEQW--INFNDT---TYVYRIV-----LRP 223
DB 745 -----SIDDLLIAEGNAYGDAFTNGVYMOTLFILOPVVVEPGSVSPGHIITEGSIIP 799

```

```

QY 224 TVDVPDPT--EGLVRTVSDYRLTYKAINC-----EAMNPVLVDGCFWIGGOYALTPSL 275
DB 800 PVNLPSSTTFODALOGT-----NITAKLVSSNGVIVNEANIP-LSPNGIYFG--YLIPKNT 852
QY 276 PQ-YDV-----SEAYALHTLTFARPSSAAL-----AFVWAGLPQ 309
DB 853 PSGLVNLLFATYYSYTLNTTIRGFIYGOIYVSNQATISVKSVMYAFEGQVFIYANTIN 912
QY 310 CGTA-----PAG-----TPAEQASSGGYLTWRHNGT---TFPAGS 342
DB 913 GTNEIKFGMFSAFVYVPSLSFNYTTISSIIEIPLMYPNKIG--EMEGNFTLPSAISAGN 969
QY 343 VSVVLEPGFALERYDPNNGSWTDFASAGDTVEFROVADEVVYVNTNPPAGSGSAPFEYRV 402
DB 970 LTYLAGQGYF-----GVPER-VLITGISALGNPTTNSGNAYNTIN 1009
QY 403 PPSNAYTNTVFRRNTLLETRPS 423
DB 1010 LPTLTFTN-----QLDKTLPS 1026

```

Search completed: May 29, 2002, 11:41:27  
 Job time: 159 sec

